Probabilistic Gaze Estimation Without Active Personal Calibration

Jixu Chen Qiang Ji
Department of Electrical,Computer and System Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180
chenji@ge.com qji@ecse.rpi.edu

Abstract

Existing eye gaze tracking systems typically require an explicit personal calibration process in order to estimate certain person-specific eye parameters. For natural human computer interaction, such a personal calibration is often cumbersome and unnatural. In this paper, we propose a new probabilistic eye gaze tracking system without explicit personal calibration. Unlike the traditional eye gaze tracking methods, which estimate the eye parameter deterministically, our approach estimates the probability distributions of the eye parameter and the eye gaze, by combining image saliency with the 3D eye model. By using an incremental learning framework, the subject doesn't need personal calibration before using the system. His/her eye parameter and gaze estimation can be improved gradually when he/she is naturally viewing a sequence of images on the screen. The experimental result shows that the proposed system can achieve less than three degrees accuracy for different people without calibration.

1. Introduction

Gaze tracking is the procedure of determining the point-of-gaze in the space, or the visual axis of the eye. Gaze tracking systems are primarily used in the Human Computer Interaction (HCI) and in the analysis of visual scanning pattern. In HCI, the eye gaze can serve as an advanced computer input [8] to replace the traditional input devices such as a mouse pointer [19]. Also, the graphic display on the screen can be controlled by the eye gaze interactively [20]. Since visual scanning patterns are closely related to the the person's attentional focus, cognitive scientists use the gaze tracking system to study human's cognitive processes [10, 11].

In general, the video-based eye gaze estimation algorithms can be classified into two groups: 2D mapping based gaze estimation methods [18, 14, 20, 21] and 3D gaze estimation methods [1, 4, 16, 2] which estimate the 3D visual axis of the subjects. A survey of eye tracking techniques may be found in [5]. Recently, the 3D methods are becoming more popular because of their high accuracy under free head movement. However, current advanced 3D gaze estimation systems require a calibration procedure for each subject in order to estimate his/her specific eye parameters. In this work, we propose a novel method to estimate eye gaze without explicit calibration procedure. In contrast to the traditional calibration procedure which asks the subject to fixate on several points on the screen, we track the eye gaze when the subject is naturally looking at the images on the screen. Combining image saliency with the 3D eye model, our method incrementally estimates the eye parameters and the eye gaze naturally without explicit calibration.

2. Related Work

In traditional gaze estimation methods, a 2D mapping approach learns a polynomial mapping from the 2D features, e.g. 2D pupil glint vector [14, 13, 20, 21], or 2D eye images [18] to the gaze point on the screen.

However, the 2D mapping approach has two common drawbacks. First, in order to learn the mapping function, the user has to perform a complex experiment to calibrate the parameters of the mapping functions. For example, in the calibration procedure of [7], the subject needs to gaze at nine evenly distributed points on the screen or gaze at twelve points for greater accuracy. Secondly, because the extracted 2D eye image features change significantly with head position, the gaze mapping function is very sensitive to head motion. Morimoto and Mimica [14] reported detailed data showing how the gaze tracking systems decay as the head moves away from the original calibration position. Hence, the user has to keep his head unnaturally still in order to achieve good performance. Methods have also been proposed to handle head pose changes using Neural Network [20] or SVM [21]. These methods, however, either only consider the in-plane head translation [20] or need stereo cameras to obtain the 3D eye position [21].

In contrast, the 3D gaze estimation is based on high res-

olution stereo cameras[16, 1, 4, 2] or a single camera with multiple calibrated light sources [3] to estimate 3D eye features (the corneal center, the pupil center, and the optical axis connecting them) directly by the 3D reconstruction technique. The visual axis is estimated from the 3D features, and the gaze point on the screen is obtained by intersecting the visual axis with the screen. However, this type of method still needs the person-specific calibration to estimate the eye parameters. For example, Chen et al. [2] proposed a 3D gaze estimation system with two cameras and IR light on each camera. Their method starts with the reconstruction of the optical axis of the eye. The visual axis can be estimated by adding a constant angle to the optical axis. However, the angle between the visual and optical axes needs to be estimated beforehand through a four-point personal calibration procedure. Guestrin et al. [4] proposed to estimate 3D gaze with two cameras and four IR lights. Their calibration procedure only required the subject to look at one point on the screen.

Most recently, some gaze estimation methods that don't use calibration have been suggested. Model and Eizenman [12] proposed to estimate the eye parameters based on the assumption that the visual axes of two eyes intersect on the screen. However, because of the noise in optical axis, it is difficult to achieve accurate result. For a standard 40cm$\times$30cm flat monitor, when the noise of the optical axis is one degree, the error of the visual axis is over five degrees. Although they proposed the use of a larger monitor (160cm$\times$120cm) or a pyramid observation surface to reduce the error, these devices are often not available in real applications.

Sugano et al. [17] offer a 2D appearance-based gaze estimation without calibration. They conduct experiment when the subject is watching an image or video in the monitor. Given the saliency map of the image, gaze points are sampled from it and used as training data to train a mapping function (Gaussian Process Regresser) between eye image and gaze point. However, because of the large uncertainty in saliency map, the accuracy of this system is rather low (~ 6 degrees) compared with state-of-the-art techniques [3] (< 1 degree). Furthermore, since this is a 2D mapping method which doesn't consider head pose, a chin rest must be used to fix the head.

In this paper, we propose an incremental probabilistic 3D gaze estimation method which allows free head movement and without explicit calibration. This method is based on combining the saliency map with the 3D eye model. First, unlike traditional 3D methods, which estimate eye parameter and gaze deterministically, the proposed method estimates the probability of eye parameter and eye gaze, and can better handle the uncertainty in the system. Second, we proposed an incremental learning method to improve estimation result gradually when the subject is naturally using the system. In both cases, no explicit calibration process or calibration targets are used. The experimental result shows that our system achieves less than three degrees average accuracy for different people.

3. 3D Gaze Estimation

Before introducing our method, we briefly summarize the 3D gaze estimation techniques.

3.1. 3D Eyeball structure

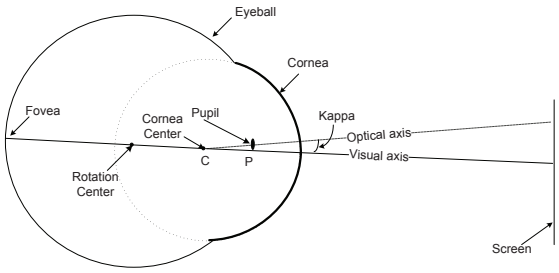

Figure 1. The structure of the eyeball.

As shown in Figure 1, the eyeball is made up of the segments of two spheres of different sizes [15]. The smaller anterior segment is the cornea. The cornea is transparent, and the pupil is inside the cornea. The optical axis of the eye is defined as the 3D line connecting the center of the pupil (**p**) and the center of the cornea (**c**). The visual axis is the 3D line connecting the corneal center (**c**) and the center of the fovea (i.e. the highest acuity region of the retina). Since the gaze point is defined as the intersection of the visual axis rather than the optical axis with the scene, the relationship between these two axes has to be modeled. The angle between the optical axis and visual axis is named *kappa* (κ), which is a constant value for each person. In traditional methods, κ is estimated through a personal calibration.

3.2. Personal Calibration

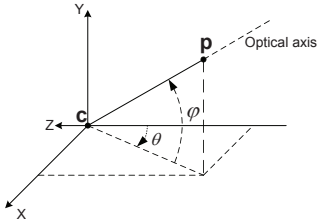

Figure 2. The orientation of optical axis.

Here, we implement the 3D gaze estimation system in [4, 3], where the cornea center **c** and optical axis **o** are directly estimated from a single camera and multiple infrared

lights. The image resolution of our camera is 640×480 pixels and the eye size in the image is about 120×60 pixels. The estimated 3D optical axis can be represented by horizonal and vertical angles ($\mathbf{o} = (\theta, \varphi)$) as shown in Figure. 2. The unit vector of the optical axis is represented as:

$$\mathbf{v}_o = \begin{pmatrix} \cos(\varphi)\sin(\theta) \\ \sin(\varphi) \\ -\cos(\varphi)\cos(\theta) \end{pmatrix} \quad (1)$$

The subject's visual axis is estimated by adding $\kappa = (\alpha, \beta)$ to the optical axis:

$$\mathbf{v}_g = \begin{pmatrix} \cos(\varphi+\beta)\sin(\theta+\alpha) \\ \sin(\varphi+\beta) \\ -\cos(\varphi+\beta)\cos(\theta+\alpha) \end{pmatrix} \quad (2)$$

Finally, the gaze point $\mathbf{g}$ on the screen is estimated by intersecting $\mathbf{v}_g$ with the screen. Thus, it is determined by the optical axis and κ :

$$\mathbf{g} = \mathbf{g}(\mathbf{o}, \kappa). \quad (3)$$

However, because κ varies for different subjects, it needs to be estimated beforehand through calibration. In traditional methods [3, 2, 4], the subject is asked to look at N specific calibration points on the screen: $\mathbf{g}_i^*, i = 1, .., N$. The eye parameter can then be estimated by minimizing the distance between the estimated gaze points and these ground-truth gaze points:

$$\kappa^* = \arg\min_{\kappa} \sum_i \|\mathbf{g}_i^* - \mathbf{g}(\mathbf{o}_i^*, \kappa)\| \quad (4)$$

where $\mathbf{o}_i^*$ is the optical axis when the subject is looking at the ith gaze point $\mathbf{g}_i^*$. The traditional gaze estimation method can be represented as Figure 3

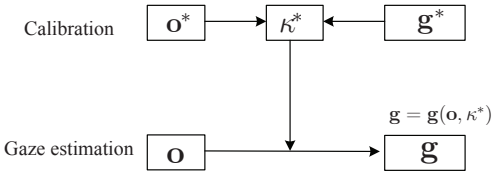

Figure 3. Diagram of traditional 3D gaze estimation. where $\mathbf{g}^*$ is ground-truth gaze.

4. Probabilistic Gaze Estimation

In the traditional methods, in order to acquire the ground-truth gaze points to estimate κ, the subject has to look at some specific points. This procedure is often cumbersome and unnatural. In this paper, we propose a new framework to estimate the probability of κ and eye gaze without forcing the subject to looking at specific calibration points.

4.1. Proposed Probabilistic Framework

The basic idea is to combine the 3D gaze estimation method with a visual saliency map. In our experiment, the subject naturally viewed the images on the screen (in full-screen mode). We utilized the method in [6] to estimate the saliency map of each image, which represents the distinctive features in the image. The only assumption we have is that the user has a higher probability of looking at the salient regions of the image. Some examples of saliency maps are shown in Figure 4. The experimental results in [6] shows

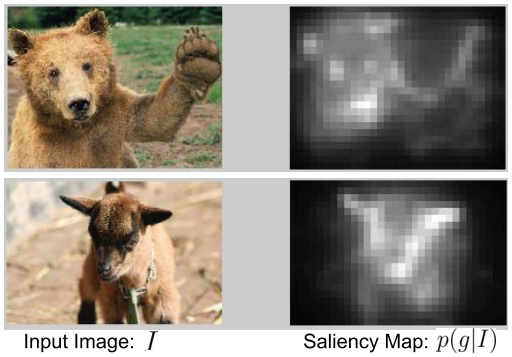

Figure 4. Examples of saliency map ($p(g|I)$)

remarkable consistency between the saliency map and the gaze. Thus, given the image (I)on the screen, its saliency map can be represented as the conditional probability of the gaze position $p(\mathbf{g}|I)$.

Based on this gaze probability, we propose the new gaze estimation framework shown in Figure 5. Notice the dif-

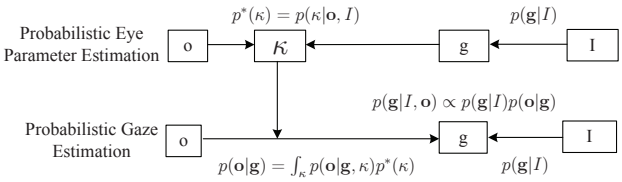

Figure 5. Diagram of the proposed probabilistic gaze estimation

ferences between our method and the traditional method in Figure 3:

1. Firstly, the traditional method needs to collect the ground-truth gaze ($\mathbf{g}^*$), when the subject is looking at specific points during calibration, while our method only needs the gaze probability $p(\mathbf{g}|I)$, when the subject is viewing the image I.

2. Secondly, the traditional method estimates the eye parameter κ^* deterministically. However, without ground-truth gaze, we cannot estimate the value of κ^*. Instead, our method estimates the probability distribution of κ : $p^*(\kappa)$.

3. Thirdly, the traditional method estimates gaze only from the optical axis and κ^*, while our method first estimates the gaze likelihood $p(\mathbf{o}|\mathbf{g})$ from the optical axis and $p^*(\kappa)$, then combines it with the gaze prior probability $p(\mathbf{g}|I)$ from the saliency map to estimate the gaze posterior probability.

This framework is mainly composed of two parts: probabilistic eye parameter estimation and probabilistic gaze estimation. We discuss them separately in the following two sections.

4.2. Probabilistic Eye Parameter Estimation

In this section, we discuss the method to estimate eye parameter (κ) probability from gaze probability (saliency map). Firstly, we introduce a general graphical model to represent the relationships between the shown image (I), eye gaze ($\mathbf{g}$), optical axis ($\mathbf{o}$), and the eye parameters (κ).

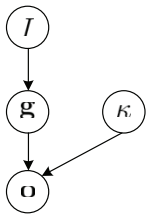

Figure 6. Probabilistic relationships in BN.

Figure 6 is the Bayesian Network (BN) [9] that represents the probabilistic relationships. The nodes in BN represent random variables, and the links represent the conditional probabilities (CPDs) of nodes given their parents. Based on the saliency map and the eye model, we define the CPDs as follows:

1. $p(\mathbf{g}|I)$: $\mathbf{g}$ is a two dimensional vector $\mathbf{g} = (x, y)$, which represents the location of the gaze on the screen (Based on the resolution of the monitor, the gaze position is discrete in the range: $0 < x < 1280, 0 < y < 1024$. The link $I \rightarrow \mathbf{g}$ is quantified by $p(\mathbf{g}|I)$ which is the saliency map estimated from image.

2. $p(\mathbf{o}|\mathbf{g}, \kappa)$: $\mathbf{o}$ has two parents $\mathbf{g}$ and κ. As discussed above, the camera in a gaze system cannot directly observe the visual axis and gaze. It can only observe the optical axis ($\mathbf{o}$) as the measurement of gaze ($\mathbf{g}$). In the traditional method, $\mathbf{o}$ is a deterministic function of $\mathbf{g}$ by subtracting a constant bias κ. In our proposed method, considering the noise in the gaze system, we model the conditional probability as a Gaussian distribution:

$$p(\mathbf{o}|\mathbf{g}, \kappa) = \mathcal{N}(f(\mathbf{g}, \kappa), \Sigma) \quad (5)$$

Here, $\mathbf{o} = (\theta, \varphi)$ is a two-dimensional vector. $f(\mathbf{g}, \kappa)$ is the inverse function of Eq.3, which estimates the optical axis by subtracting κ from the visual axis. Σ models the noise in the gaze tracking system, which is estimated from training data. (According to the tests of our system, we set the standard deviation of the optical axis as one degree on both θ and φ.)

Now, based on the BN model, eye parameter estimation is solved as an inference problem in the BN, which estimates the posterior probability $p(\kappa|\mathbf{o}, I)$ given the optical axis and the shown image. Based on the conditional independencies in the BN model, the probability of κ can be written as:

$$p(\kappa|\mathbf{o}, I) \propto \int_{\mathbf{g}} p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g}, \kappa) \cdot p(\kappa) \propto \int_{\mathbf{g}} p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g}, \kappa) \quad (6)$$

$p(\mathbf{g}|I)$ is the saliency map; $p(\mathbf{o}|\mathbf{g}, \kappa)$ is the Gaussian distribution as defined in Eq.5. The prior probability of κ ($p(\kappa)$) is initially assumed to be a uniform distribution. Here, Eq.6 is a one-step belief propagation that propagates the probability from the gaze to κ given one optical axis. The gaze position is discrete in a limited range; thus, the integral in the above equation can be approximated by summation.

Fig. 7(C) shows an example of the estimated eye parameter probability. Here, we collected 40 optical axes when the subject was looking the image in Fig. 7(A). Thus, the training optical axes are $\mathbf{o}_{1,..,40}$ and their corresponding shown images $I_{1,..,40}$ are the same. Assuming these optical axes are conditionally independent to each other, we can derive the κ probability as the product of each single probability:

$$p(\kappa|\mathbf{o}_{1,..,40}, I_{1,..,40}) \propto \prod_{i=1}^{40} \int_{\mathbf{g}_i} p(\mathbf{g}_i|I_i)p(\mathbf{o}_i|\mathbf{g}_i, \kappa) \quad (7)$$

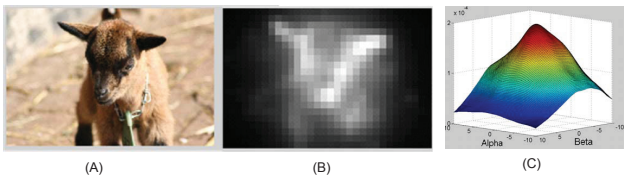

Figure 7. Probabilistic Eye Parameter Estimation. (A) is the shown image. (B) is the saliency map $p(\mathbf{g}|I)$ of the image. (C) is the estimated probability distribution of eye parameter $p^*(\kappa)$. (x-axis represents α and y-axis represents β.)

Based on the biological study, eye parameters should be in a limited range for normal eyes. Here we restricted the eye parameter in the range $-10^o < \alpha < 10^o$ and $-10^o < \beta < 10^o$.

4.3. Probabilistic Gaze Estimation

Given the estimated eye parameter probability $p^*(\kappa)$, we can estimate the gaze probability. For consistency, this

derivation is based on the same BN model in Fig.6. Unlike the eye parameter estimation, the estimated $p^*(\kappa)$ is now used as the prior probability of the κ node. Then, the probability of the gaze, the optical axis and the shown image, can be written:

$$p(\mathbf{g}|\mathbf{o}, I) \propto p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g}) \tag{8}$$

where $p(\mathbf{g}|I)$ is the prior probability of gaze from the saliency map of the shown image I, and $p(\mathbf{o}|\mathbf{g})$ is the gaze likelihood, which can be derived from $p^*(\kappa)$ as:

$$p(\mathbf{o}|\mathbf{g}) = \int_{\kappa} p(\mathbf{o}|\mathbf{g}, \kappa)p^*(\kappa) \tag{9}$$

Note that all the above derivations are only valid based on the conditional independencies in the BN model.

Thus, the probabilistic gaze estimation is composed of the following steps:

1. Firstly, we estimated the gaze prior probability from the saliency map $p(\mathbf{g}|I)$.
2. Then, we estimated the likelihood gaze map $p(\mathbf{o}|\mathbf{g})$, given the current optical axis and the eye parameter prior $p^*(\kappa)$.
3. Finally, the product $p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g})$ represents gaze posterior probability map. The maximum posterior point is selected as the gaze point.

The results of the three steps are shown in Fig.8. Here we compare our method with the traditional gaze estimation method, which uses 9-point calibration to determine the eye parameter. The peak in our posterior probability map is very close to the estimated gaze of the traditional method, but our method does not need any explicit calibration.

4.4. Incremental Learning for Gaze Estimation

The above probabilistic framework includes two stages: first, $p^*(\kappa)$ is estimated when the subject is looking at the training images. Then his/her gaze is estimated when he/she is looking at the test images.

In order to provide a more natural user experience, we propose an incremental learning algorithm for our probabilistic framework. This new framework does not need any prior training. It can quickly adapt to the user, and incrementally improves gaze estimation accuracy as the subject uses the system.

We first assume the initial distribution of κ as uniform. When the subject starts to use the system, we record a sequence of his optical axes $\mathbf{o}_{t,..,1}$. Given the corresponding shown image sequence $I_{t,..,1}$, the incremental learning framework continually updates the estimations of κ and gaze given all previous information, i.e. estimating $p(\kappa_t|I_{t,..,1}, \mathbf{o}_{t,..,1})$ and $p(\mathbf{g}_t|I_{t,..,1}, \mathbf{o}_{t,..,1})$. We employ a recursive updating procedure detailed as follows.

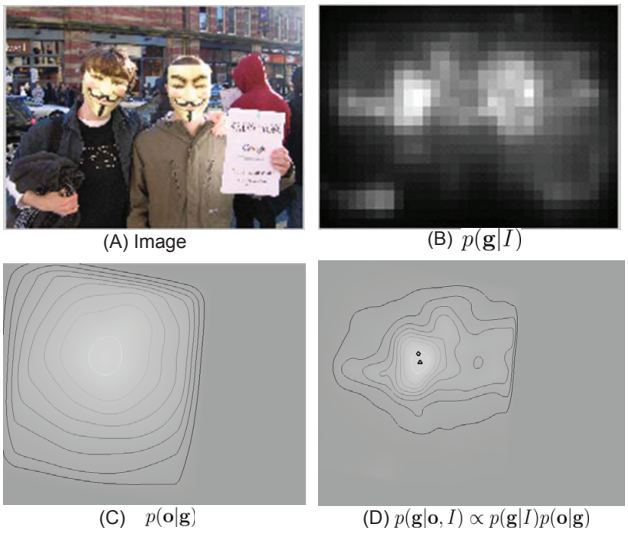

Figure 8. Probabilistic Gaze Estimation. (A) is the shown image. (B) is the saliency map $p(\mathbf{g}|I)$ of the image. (C) is gaze likelihood map given optical axis. (D) is the gaze posterior probability map. The black triangle shows the maximum posterior point. The circle shows the estimated gaze using traditional method.

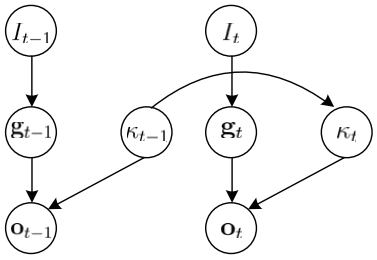

Figure 9. DBN for incremental learning.

For incremental learning, we first extend the BN to a dynamic BN (DBN) model as shown in Figure 9. The DBN includes two kind of links. *intra-frame links* in one time frame are the same as the BN model, and *inter-frame link* from κ_{t-1} to κ_t capture the temporal relationships. Base on the anatomy, κ cannot vary much over time. Thus, we model it as a Gaussian distribution:

$$p(\kappa_t|\kappa_{t-1}) = \mathcal{N}(\kappa_{t-1}, \Sigma_k) \tag{10}$$

where Σ_k is the covariance matrix which allows κ_t to vary in a small range around the previous estimation κ_{t-1}. Here, according to the previous user study, we set the standard deviations of α_t and β_t to one degree, i.e. Σ_k is an identity matrix.

Given the above temporal relationship, the probability of κ can be updated recursively. Firstly, we predicted the prior probability of the current κ_t based on its previous probabil-

ity, as shown in Eq. 11.

$$p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) = \int_{\kappa_{t-1}} p(\kappa_t|\kappa_{t-1}) p(\kappa_{t-1}|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \quad (11)$$

where $p(\kappa_{t-1}|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$ is the κ probability from the previous time frame. Since the temporal CPD $p(\kappa_t|\kappa_{t-1})$ is a Gaussian distribution, this integral is implemented by a convolution of previous κ probability map with a Gaussian kernel. In the first time frame, when no prior information of κ is available, we assume $p(\kappa_1)$ is uniformly distributed.

Based on the predicted temporal prior probability of κ_t, the current probability of $\mathbf{g}_t$ and κ_t can be derived as the filtering problem in the DBN:

$$p(\mathbf{g}_t|I_{t,t-1,..,1}, \mathbf{o}_{t,t-1,..,1}) \propto p(\mathbf{g}_t|I_t) \cdot \int_{\kappa_t} p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \quad (12)$$

$$p(\kappa_t|I_{t,t-1,..,1}, \mathbf{o}_{t,t-1,..,1}) \propto \int_{\mathbf{g}_t} p(\mathbf{g}_t|I_t) p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) \cdot p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \quad (13)$$

Let $p^*(\kappa_t) = p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$ and $p'(\kappa_t) = p(\kappa_t|I_{t,..,1}, \mathbf{o}_{t,..,1})$, the above incremental learning algorithm can be summarized as follows:

Algorithm 2 Incremental Gaze Estimation Algorithm

$t \leftarrow 1$

Set $p^*(\kappa_1)$ as uniform distribution.

Estimate the first gaze:
$p(\mathbf{g}_1|I_1, \mathbf{o}_1) \propto p(\mathbf{g}_1|I_1) \cdot \int_{\kappa_1} p(\mathbf{o}_1|\mathbf{g}_1, \kappa_1) p^*(\kappa_1)$

Update κ probability :
$p'(\kappa_1) \propto \int_{\mathbf{g}_1} p(\mathbf{g}_1|I_1) p(\mathbf{o}_1|\mathbf{g}_1, \kappa_t)$

loop

$t \leftarrow t+1$

Temporal belief propagation $p'(\kappa_{t-1}) \rightarrow p^*(\kappa_t)$:
$p^*(\kappa_t) = \int_{\kappa_{t-1}} p(\kappa_t|\kappa_{t-1}) p'(\kappa_{t-1})$

Probabilistic gaze estimation:
$p(\mathbf{g}_t|I_{t,..,1}, \mathbf{o}_{t,..,1}) \propto p(\mathbf{g}_t|I_t) \cdot \int_{\kappa_t} p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) p^*(\kappa_t)$

Update κ probability:
$p'(\kappa_t) \propto \int_{\mathbf{g}_t} p(\mathbf{g}_t|I_t) p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) \cdot p^*(\kappa_t)$

end loop

The only difference between the DBN and the BN is that the DBN considers the temporal prior of κ_t, and continues updating it over time. For example, if letting $p^*(\kappa) = p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$, Eq. 12 is the same as Eq. 8; if letting $p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$ be uniform distribution, Eq. 13 is the same as Eq. 6.

An example of the incremental learning of $p'(\kappa_t)$ is shown in Figure 10. The estimated $p'(\kappa_1)$ for the first time frame has a high probability in multiple regions. By updating its probability incrementally, it gradually converges to a single peak after twenty time frames.

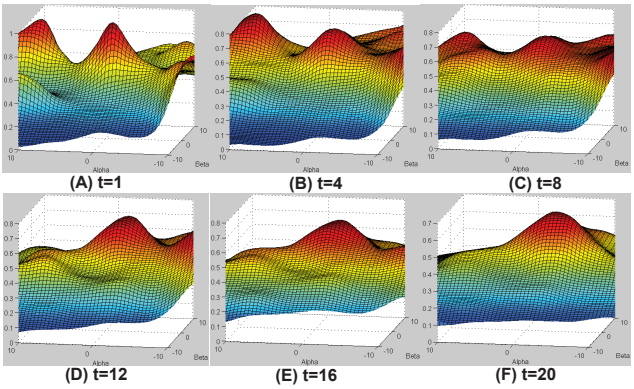

Figure 10. Incremental learning of $p'(\kappa_t)$.

5. Experimental Results

To evaluate the traditional gaze estimation method, the subjects are often asked to look at some points on the screen. The gaze estimation error can be computed as the distance between these points and the estimated gaze points. However, in our method, the user does not need to look at any specific points. To evaluate our system, we implement the traditional 3D gaze estimation system [3]. This system is first calibrated by asking the subject to look at nine points on the screen. The average accuracy of this system is one degree for different subjects. We compared our proposed method with this system.

To evaluate our method, we collected the optical axes of five subjects while they viewed the images on the screen. Each image displayed for 3-4 seconds on the screen. We collected 80 optical axes for each image (our gaze system captured the video of the eye and estimated the optical axes at 25 frames per second). To show the advantages of the incremental learning, we compared the incremental learning algorithm in Section 4.4 to the batch training method in Section.4.2.

5.1. Batch Training for Gaze Estimation

For batch training, we divided the 80 time frames when the subject viewed one image into training data (40 frames) and testing data (40 frames). Each subject viewed five images in this experiment, and we used leave-one-image-out cross validation, i.e. when testing on 40 frames of one image, we first learned the eye parameter probability from the training data of the other four images (Section.4.2).

For a more effective system, we want to use less training data, because more training time may make the subject

bored and easily distracted. We tested the dependency on the training data by using 160, 80, 40, and 20 frames of training data, i.e. 40, 20, 10, and 5 frames for each training image.

The average error (over 200 test frames) of each subject is show in Table 1. Performance did not decrease much when training frames were reduced to 40. Further reducing the training data resulted in an increase of error. The average gaze estimation error of our proposed method achieved 2.40^o when there was enough training data (160 frames).

5.2. Incremental Learning for Gaze Estimation

Based on our incremental learning algorithm, the system doesn't need to estimate the eye parameter probability beforehand using training frames. This system can automatically update the eye parameter probability and estimate the gaze when the subject start using the system.

The gaze estimation error for the first 10, 20, 40, 80, 120, 160, and 200 frames are shown in Table 2. Although the error is large for the first few frames (<20 frames), it decreases quickly as the subject uses the system. Compared with the batch training, the incremental learning achieves similar performance for the first twenty frames. However, when the subject is using the system, incremental learning keeps improving the performance and can achieve an average accuracy of 17.07mm (1.77^o) for the first 200 frames. This process is done automatically, naturally, and without any user knowledge.

Some gaze estimation results (in both the original image and the saliency map) of subject 1 are shown in Figure 11. Without calibration, the results of our method are close to the results of the system with 9-point calibration. The subject may look at some region with low saliency, such as the white paper in the person's hand in Figure 11(A). In this case, by incrementally improving the eye parameter estimation and by combining gaze likelihood with the saliency map, our method can still follow the true gaze positions.

Compared to the most recent calibration-free gaze estimation method [17], which asks the subject to watch a ten-minute video for training and achieves an accuracy of six degree, our proposed method doesn't need training data beforehand and can adapt to the user very quickly (in 80 frames or three seconds), and continues to improve the accuracy when person start using it. The average accuracy can achieve 1.77 degrees. Furthermore, in our 3D gaze estimation framework, the subject can make natural head movement without a chin-rest.

6. Conclusion

In this paper, we proposed a new probabilistic gaze estimation framework by combining the saliency map with the 3D eye gaze model. Compared to the traditional method, our proposed approach doesn't need the cumbersome and unnatural personal calibration procedure. Compared with the most recent calibration-free method [17], our system allows natural head movement. In addition, by considering the uncertainties of eye parameter and gaze in our probabilistic framework, our system significantly improves the accuracy from six degrees to less than three degrees. By using a novel incremental learning framework, our system doesn't need any training data from the subject beforehand. It can adapt to the user quickly and improves its performance as the subject naturally using the system.

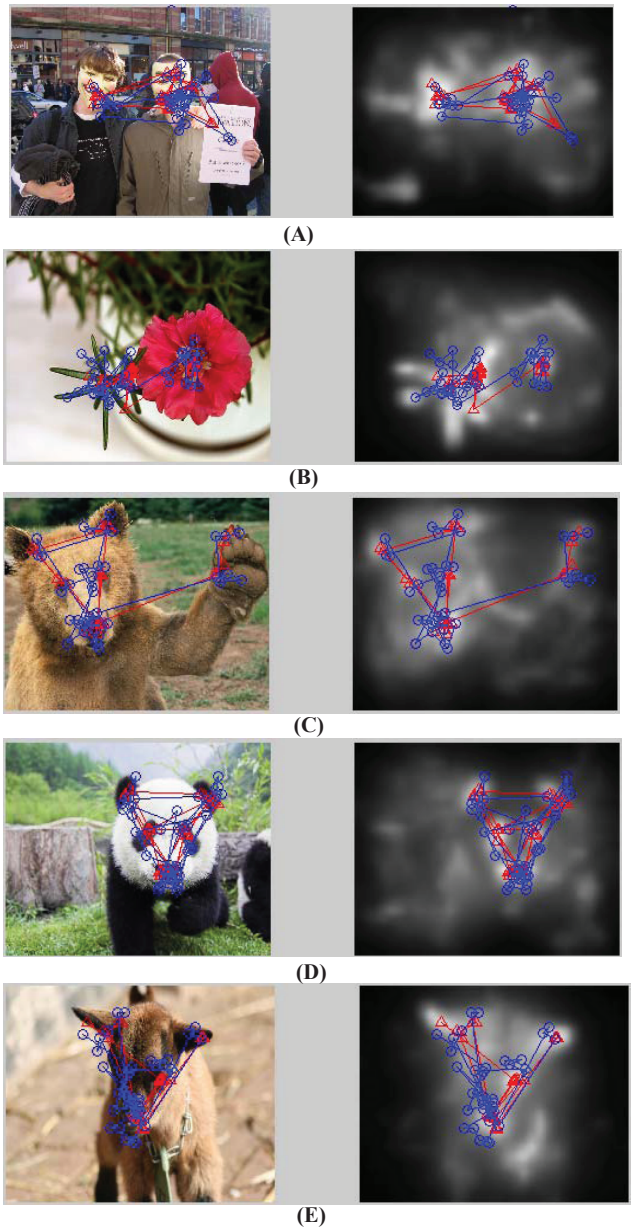

Figure 11. Probabilistic Gaze Estimation Result. The red rectangles are the results of our proposed method. The blue circles are the results of traditional method with 9-point calibration.

Table 1. Gaze estimation error of 5 subjects with different training data size. Eye parameters are trained though batch training.

Training data size	160 frames		80 frames		40 frames		20 frames	
	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]
Subject 1	21.00	2.18	23.13	2.40	23.36	2.43	23.59	2.45
Subject 2	18.55	1.93	18.49	1.92	17.57	1.83	19.03	1.98
Subject 3	15.47	1.61	15.76	1.64	15.99	1.66	16.38	1.70
Subject 4	27.84	2.89	26.68	2.77	28.17	2.93	28.43	2.95
Subject 5	32.73	3.40	32.15	3.34	32.61	3.39	33.04	3.43
Average	23.11	2.40	23.24	2.41	23.54	2.45	24.09	2.50

Table 2. Gaze estimation results of 5 subjects for the first N frames (N=10,20,40,80,120,160,200). Eye parameters are automatically updated after each frame.

	10 frames		20 frames		40 frames		80 frames		120 frames		160 frames		200 frames	
	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]
Subject 1	23.57	2.45	16.66	1.73	19.25	2.01	19.89	2.07	19.64	2.04	18.20	1.89	17.32	1.80
Subject 2	32.91	3.42	24.03	2.50	19.79	2.06	18.32	1.90	17.73	1.84	18.20	1.89	17.33	1.80
Subject 3	16.46	1.71	15.42	1.60	15.16	1.57	14.85	1.54	14.48	1.50	15.36	1.59	15.36	1.59
Subject 4	26.57	2.76	39.36	4.09	28.42	2.95	21.18	2.20	18.41	1.91	19.85	2.06	19.92	2.07
Subject 5	28.41	2.95	28.34	2.95	27.07	2.81	20.11	2.09	16.20	1.68	15.40	1.60	15.41	1.60
Average	25.58	2.66	24.76	2.57	21.94	2.28	18.87	1.96	17.29	1.80	17.40	1.81	17.07	1.77

7. Acknowledgement

The authors gratefully acknowledge Prof. Moshe Eizenman of Univ. of Toronto for inspiring the idea in the paper and for many helpful discussions.

References

[1] D. Beymer and M. Flickner. Eye gaze tracking using an active stereo head. *IEEE Conference in CVPR03*, 2003. 609, 610

[2] J. Chen, Y. Tong, W. Gray, and Q. Ji. A robust 3d eye gaze tracking system using noise reduction. *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 189–196, 2008. 609, 610, 611

[3] E. D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transcations on Biomedical Engineering*. 610, 611, 614

[4] E. D. Guestrin and M. Eizenman. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. *Proceedings of the 2008 symposium on Eye tracking research & applications*, 2008. 609, 610, 611

[5] D. W. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 2010. 609

[6] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *NIPS*, 2006. 611

[7] L. T. Inc. http://www.eyegaze.com. 609

[8] R. J. Jacob. The use of eye movements in human computer interaction technqiues: What you look at is what you get. *ACM Tracnsations on Information Systems*, 9:152–169, 1991. 609

[9] D. Koller and N. Friedman. *Probabilistic Graphical Models : Principles and Techniques*. The MIT Press, 2009. 612

[10] S. Liversedge and J. Findlay. Saccadic eye movements and cognition. *Trends in Cognitive Science*, 4:6–14, 2000. 609

[11] M. Mason, B.Hood, and C. Macrae. Look into my eyes : Gaze direction and person memory. *Memory*, 12:637–643, 2004. 609

[12] D. Model and M. Eizenman. An automatic personal calibration procedure for advanced gaze estimation systems. *IEEE Transactions on Biomedical Engineering*, 57(5), 2010. 610

[13] C. H. Morimoto, A. Amir, and M. Flickner. Detecting eye position and gaze from a single camera and 2 light sources. *Proceedings of the International Conference on Parttern Recognition*, 2002. 609

[14] C. H. Morimoto and M. R. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding, Special Issue on Eye Detection and Tracking*, 98:4–24, 2005. 609

[15] C. W. Oyster. *The Human Eye: Structure and Function*. Sinauer Associate, Inc., 1999. 610

[16] S.-W. Shih and J. Liu. A novel approach to 3-d gaze tracking using stereo cameras. *IEEE Transactions on Systems, Man and Cybernetics, PartB*, 34:234–245, 2004. 609, 610

[17] Y. Sugano, Y. Matsushita, and Y. Sato. Calibration-free gaze sensing using saliency maps. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 610, 615

[18] K. Tan, D. Kriegman, and H. Ahuja. Appearance based eye gaze estimation. *Proceedings of IEEE Workshop Applications of Computer Vision*, pages 131–136, 2002. 609

[19] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 246–253, 1999. 609

[20] Z. Zhu and Q. Ji. Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications*, 15:139–148, 2004. 609

[21] Z. Zhu, Q. Ji, and K. P. Bennett. Nonlinear eye gaze mapping function estimation via support vector regression. *International Conference on Pattern Recognition*, 2006. 609